Digital Audio Watermarking Based on Time-Spread Echo Hiding(エコー拡散法に基づくディジタル音信号電子透かしの研究)

著者	高 秉燮
号	288
発行年	2003
URL	http://hdl.handle.net/10097/12985

氏名（本籍）	KO BYEONG SEOB 高 秉 燮 (韓国)
学位の種類	博士 (情報科学)
学位記番号	情博第 288 号
学位授与年月日	平成 16 年 3 月 25 日
学位授与の要件	学位規則第 4 条第 1 項該当
研究科，専攻	大学院情報科学研究科 (博士課程) システム情報科学専攻
学位論文題目	Digital Audio Watermarking Based on Time-Spread Echo Hiding (エコー拡散法に基づくディジタル音信号電子透かしの研究)
論文審査委員	(主査) 東北大学教授 鈴木 陽一　東北大学教授 静谷 啓樹 東北大学教授 牧野 正三 (工学研究科)

論 文 内 容 要 旨

Chapter 1 Introduction

The motivation of this study stems from protecting copyright of digital media contents from piracy. The progress of personal computers and digitalized equipments such as digital recording/storage devices has make it possible to easily create, replicate, and edit digital music, image, and movie contents. Such digital media contents can be rapidly shared, transmitted, and distributed via the broad-band networks like Internet. It brought about several social problems such as violation of copyright and abuse of digital media contents. Thus, it is regarded as a very serious problem for wholesome progression of the industry of digital media contents, the IT industry, and the global-network society. Accordingly, digital watermarking techniques has received much more attention for copyright protection of digital media contents. Many techniques have been proposed for watermarking of digital audio contents over the last few decade. However, the techniques have several demerits which have to be solved for practical applications.

The goal of this dissertation is, therefore, to propose a novel digital audio watermarking technique for copyright protection for digital audio contents. To make it applicable for the practical applications, the improvement and evaluation of the proposed method are carried out by computer simulations and listening tests.

Chapter 2 Time-Spread Echo Hiding Using PN Sequences

Conventional watermarking techniques based on echo hiding provide many benefits such as simple embedding/decoding process and good robustness, but also have several disadvantages such as very low secrecy because the decoding process is very lenient. In this chapter, a novel time-spread echo as an alternative to the single echo in conventional echo hiding is proposed to solve the weak points of conventional echo hiding. The concept of a spread-spectrum technique is exploited to spread the echo in the time domain.

Spreading the echo of conventional echo hiding is achieved by pseudo-noise (PN) sequences which are generally used in spread-spectrum techniques to spread data. To assure imperceptibility, temporal masking in the human auditory system (HAS) and the concept of reverberation in room acoustics are also exploited

in the spreading process. By spreading the echo using a PN sequence, the number of echoes increases up to the length of the PN sequence. However, imperceptibility can be assured because the amplitude of each echo can be substantially reduced by the spreading.

Figure 1 illustrates impulse responses of a single echo kernel and a time-spread echo kernel. In the proposed method, a single echo is temporally spread in the time domain using a PN sequence, which acts as a secret key to decode the embedded watermarks from a watermarked signal. In Fig. 1, g is the amplitude of the echo, α is the amplitude of the PN sequence, L_{PN} is the length of the PN sequence, and Δ is the time delay.

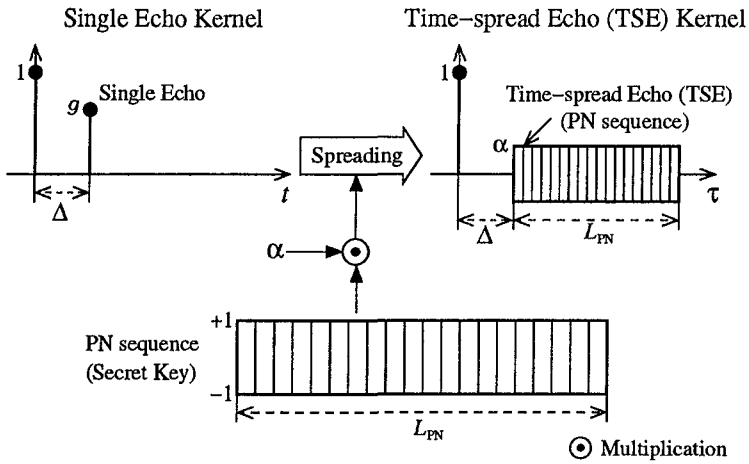

図 1: Single echo kernel and time-spread echo kernel.

The proposed kernel in Fig. 1 is constructed using a PN sequence as

$$k(n) = \delta(n) + \alpha \cdot p(n - \Delta), \quad 0 < \alpha \ll 1, \tag{1}$$

where $p(n)$ is an original PN sequence whose amplitude is ± 1, α is the amplitude of PN sequence, and $\delta(n)$ is the Dirac delta function.

The watermarked signal is obtained by taking a linear convolution of the host signal and the time-spread echo kernel. Thus, the watermarked signal is denoted as

$$w(n) = s(n) * k(n), \tag{2}$$

where $s(n)$ is the host signal, $k(n)$ is the time-spread echo kernel, and $*$ is a linear convolution.

Decoding the embedded watermarks is realized by the cepstrum analysis of the watermarked signal and the cross correlation between the cepstrum and the PN sequence used in the embedding process. Thus, the decoded signal is obtained by the following process:

$$d(n) = \mathrm{F}^{-1}[\log(\mathrm{F}[w(n)])] \otimes p(n), \tag{3}$$

where $\mathrm{F}[\cdot]$ is the Fourier transform, $\mathrm{F}^{-1}[\cdot]$ is the inverse Fourier transform, $\log(\cdot)$ is the logarithm operation, and $\otimes$ means the cross-correlation operation.

Chapter 3 Robust Embedding Process for the Time-Spread Echo Hiding

The time-spread echo hiding faces a dilemma similar to that of conventional echo hiding. That is, pseudo-noise (PN) sequences with large amplitude should be employed for good decoding performance and robustness though it could spoil the sound quality of the watermarked signal. In this chapter, a new robust embedding process for the time-spread echo hiding is proposed to cope with the above-mentioned dilemma. The proposed embedding is achieved by decomposing a host signal into multiple subband signals and controlling the amplitude of the PN sequence for each subband as a function of the masking level, which is given by the masking in the frequency domain, for each subband.

Figure 2 shows a block diagram of the proposed embedding process with subband decomposition. In Fig. 2, the time-spread echo kernel for each subband, $k_i(n)$, is expressed as:

$$k_i(n) = \delta(n) + \alpha_i \cdot p(n - \Delta), \tag{4}$$

where $i = 0, 1, \cdots, M-1$ and M is the number of subbands.

A watermarked signal is obtained as follows:

1. Dividing a host signal into M subband signals by M-channel filter banks with equal bandwidth.
2. Calculating the SMR of each subband from the MPEG psychoacoustic model.
3. Determining the amplitude of PN sequence, α_i, for each subband from a certain function which is defined as a function of the SMR.
4. Embedding the same watermark into each subband by convolving the corresponding time-spread echo kernel.
5. Summing all processed subband signals up to the watermarked signal, $w(n)$.

Thus, the watermarked signal, $w(n)$, can be expressed as follows:

$$\begin{aligned} & w(n) \\ &= \sum_{i=0}^{M-1} w_i(n) = \sum_{i=0}^{M-1} s_i(n) * k_i(n) \\ &= s(n) + (\alpha_0 \cdot s_0(n) + \alpha_1 \cdot s_1(n) + \cdots + \alpha_{M-1} \cdot s_{M-1}(n)) * p(n - \Delta) \\ & \quad (\text{if, } \beta \approx \alpha_0 \approx \alpha_1 \approx \cdots \approx \alpha_{M-1}) \\ &\approx s(n) * (\delta(n) + \beta \cdot p(n - \Delta)), \end{aligned} \tag{5}$$

where $s(n)$ is the host signal, $s_i(n)$ $(i = 0, 1, \cdots, M-1)$ is the decomposed host signal for each subband, $k_i(n)$ $(i = 0, 1, \cdots, M-1)$ is the time-spread echo kernel for each subband, and $w_i(n)$ $(i = 0, 1, \cdots, M-1)$ is the watermarked signal for each subband. The watermark embedded by the new method can be successfully decoded by the same decoding process as in Eq. (3) without subband decomposition. This means that the complexity of the decoding process is equal to that of the original method although the complexity of the embedding process increases at least M times more than that of the original simple process.

Chapter 4 Log-Scaling Method to Improve Robustness against Pitch-Scaling

Pitch-scaling is regarded as the severest attack in typical attacks such as MP3, AAC, sampling frequency conversion, and time-scaling. In particular, the techniques which use a pseudo-noise (PN) sequence as a secret key in the frequency or the quefrency domain are very vulnerable to pitch-scaling because of the desynchronization between the original and the embedded PN sequences.

The term "pitch-scaling" of an audio signal commonly indicates that the pitch of the signal is scaled by a scale factor β. The effect of pitch-scaling in the quefrency domain appears as follows:

$$\tau' = \beta \cdot \tau, \tag{6}$$

where β is the pitch-scaling factor in the quefrency domain, τ is the original quefrency, and τ' is the scaled quefrency. A logarithm operation is widely exploited to convert multiplication into addition. Similarly, it converts the scaling (multiplying) process into the shifting (addition) process on the logarithmic axis. That is, taking the logarithm of Eq. (6), the equation is rewritten as follows:

$$\log(\tau') = \log(\tau) + \log(\beta), \tag{7}$$

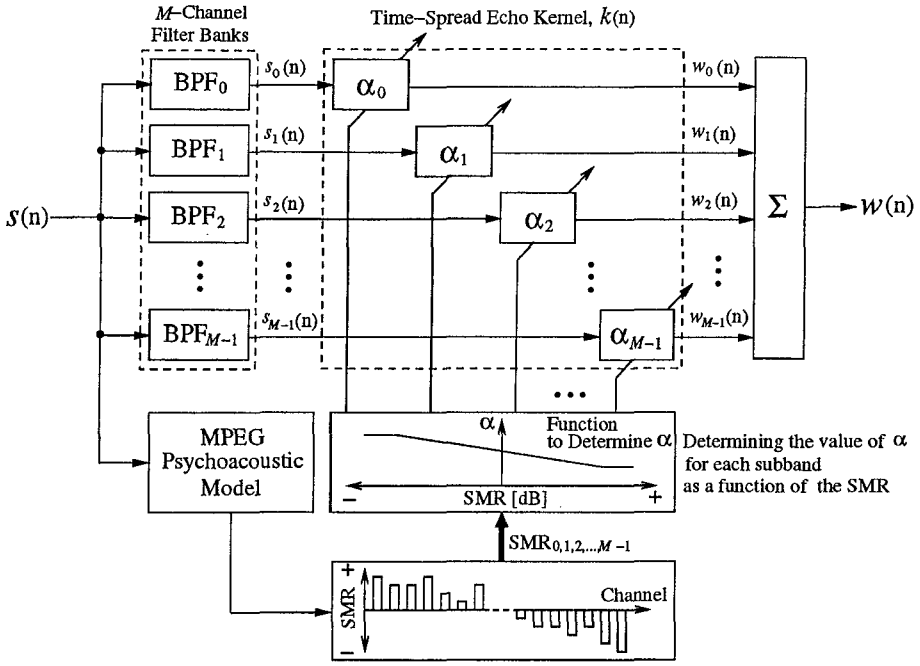

図 2: Block diagram of the proposed embedding process with subband decomposition process.

where $\log(\cdot)$ is the logarithm operation. In Eq. (7), we see that scaling by β is converted into shifting by $\log(\beta)$. This logarithm operation for the frequency or quefrency axis is referred to as "log-scaling" in this stduy. If the original version, i.e., a signal without pitch-scaling, and the pitch-scaled version are both converted by log-scaling, the patterns of the two log-scaled versions becomes accordant with each other though the pitch-scaled version is shifted according to the amount of pitch-scaling. Thus, correlation between the original and pitch-scaled version can be recovered regardless of whether there is pitch-scaling or not.

However, the log-scaling in Eq. (7) cannot be directly implemented because discrete-time signal processing is basically used in the embedding and decoding process. Thus, the log-scaling is firstly scaled as follows:

$$\tau'_L = \lfloor \gamma \cdot \log(\tau') \rfloor, \tag{8}$$

where $\lfloor \cdot \rfloor$ rounds the element to the nearest integer towards minus infinity and γ is a log-scaling factor to obtain the discrete points. To obtain the complete log-scaled version of the signal, a linear interpolation is performed to obtain the samples between the log-scaled samples.

Chapter 5 Conclusions

The goal of this dissertation has been achieved by the time-spread echo hiding proposed in Chapter 2, which was proposed to cope with the major drawback of conventional echo hiding, the robust embedding for time-spread echo hiding proposed in Chapter 3, which gave the better robustness than that of the original embedding in Chapter 2 while maintaining imperceptibility, and the log-scaling method proposed in Chapter 4, which succeeded to enhance the robustness against pitch-scaling.

The techniques proposed in this study provide imperceptibility, robustness, and secrecy enough for practical applications. The concept of the proposed methods can be extended to other digital watermarking techniques. Moreover, the log-scaling method should be useful to solve the synchronization problems came from by using PN sequences in other fields such as digital communications. Therefore, the proposed methods can be expected to be contributed to protect copyrights of sound media contents in practical applications and to solve the problems in other IT technologies.

論文審査の結果の要旨

近年，情報通信技術の発展により，映画や音楽など，ディジタルマルチメディアコンテンツの流通が拡大している．これに伴って，それらの不法コピー流通が増加し，社会的な問題の一つになっていることから，ディジタルマルチメディアコンテンツのコピー防止及び著作権保護技術が強く求められている．このための技術として，実際に視聴できる信号の状態で著作権保護が可能な「電子透かし」技術が注目されている．ディジタル音信号についても，過去いくつかの技法が提案されてきたが，透かしが埋め込まれた信号の音質，攻撃に対する耐性，セキュリティなどに問題があり，実用性に乏しかった．そこで，著者は実用性に優れた高性能な音信号用電子透かしの研究を行った．本論文は，その研究成果をまとめたもので，全編5章からなる．

第1章は序論であり，本研究の背景と目的を述べている．

第2章では，従来法のうちでは最も優れた技法と考えられるエコー法に着目し，その欠点を解決した新たな電子透かし埋め込み・検出法である「エコー拡散法」の提案を行っている．この方法では，エコー成分をPN(擬似雑音)系列を用いて時間領域で拡散させることにより，音質への影響を軽減するとともに，攻撃に対する耐性とセキュリティの向上を図っている．聴取実験と計算機シミュレーションの結果，提案技法が，音質と攻撃耐性，セキュリティの面で優れた特性を持つことを検証している．これは，実用性の高い技法の提案として評価できる．

第3章では，第2章で提案した技法の耐性を向上させるため，聴覚の特性を利用した透かし埋め込み法の改善法を提案している．ホスト信号を複数のサブバンドに分解し，それぞれのサブバンドに埋め込まれる透かし，すなわちPN系列のパワーをMPEG心理音響モデルから得られるマスキングレベルにしたがって変化させ，埋め込みを行う．これにより，第2章の提案法より大きいレベルで埋め込みを行っても音質を保つことができ，耐性を向上させることに成功した．

第4章では，ピッチ変換によりPN系列の同期が取れなくなった信号から透かしを検出するために，ケプストラム分析の後，ケフレンシー軸の対数変換を行って，ピッチ変換をケフレンシー軸上のシフトに変換することにより，オリジナルのPN系列と同期を取ることに成功した．これは従来の音信号電子透かしにおいて大きな問題であった，ピッチ変換攻撃に対し，耐性を大幅に強化する技法の提案として高く評価できる．

第5章は結論である．

以上要するに本論文は，ディジタル音信号を対象とした新しい高性能電子透かし技法の提案を行い，その有効性を示したもので，システム情報科学ならびに音響情報工学の発展に寄与するところが少なくない．

よって，本論文は博士(情報科学)の学位論文として合格と認める．